Toward Measurement of Situation Awareness in Autonomous Vehicles

David Sirkin, Nikolas Martelaro, Mishel Johns, Wendy Ju
Center for Design Research, Stanford University, CA USA
{sirkin, nikmart, mishel, wendyju}@stanford.edu

ABSTRACT
Until vehicles are fully autonomous, safety, legal and ethical obligations require that drivers remain aware of the driving situation. Key decisions about whether a driver can take over when the vehicle is confused, or its capabilities are degraded, depend on understanding whether he or she is responsive and aware of external conditions. The leading techniques for measuring situation awareness in simulated environments are ill-suited to autonomous driving scenarios, and particularly to on-road testing. We have developed a technique, named Daze, to measure situation awareness through real-time, in-situ event alerts. The technique is ecologically valid: it resembles applications people use in actual driving. It is also flexible: it can be used in both simulator and on-road research settings. We performed simulator-based and on-road test deployments to (a) check that Daze could characterize drivers' awareness of their immediate environment and (b) understand practical aspects of the technique's use. Our contributions include the Daze technique, examples of collected data, and ways to analyze such data.

Author Keywords
Autonomous vehicles; situation awareness; interaction design; measurement

INTRODUCTION
It is tempting to think that someday, when we have fully autonomous vehicles, we will be able to climb into our cars and take a nap on the way to wherever we are headed as the vehicle does the driving. However, such a day is still far off. Even when we are driven by humans, such as taxi drivers, we often feel some need to supervise drivers to make sure that they understand our intentions, that they are competent drivers, and that they are not malingering or adding unnecessary waypoints along the journey. This type of oversight will generally be required in autonomous cars until they gain the capability to perform fully autonomous driving in all possible on-road situations [22]. It is hence necessary to assess the level of situation awareness that people maintain in different autonomous driving scenarios.

CHI 2017, May 06-11, 2017, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/3025453.3025822

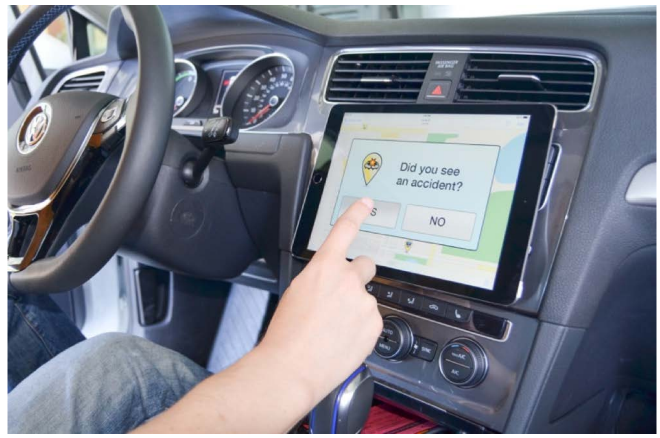

Figure 1. The Daze application, shown during on-road testing.

Through our experience developing interfaces for autonomous vehicles, we have found a need to measure situation awareness in both simulated and on-road autonomous driving settings. Some of the most commonly used techniques for assessing situation awareness depend on measures of the driver's performance, but these are inappropriate for automated driving. Other techniques [8, 9, 28, 35] require halting a simulation currently in-progress to question drivers about what they notice, but this method does not translate well for measuring situation awareness in vehicles driving on the road. The alternative to these in-situ, halting, measures is to use post-facto interviews and questionnaires, which are affected by people's recall ability and occurrences of subsequent events.

We have developed a system, *Daze*, to enable real-time, in-situ measurement of situation awareness during driving activity [17]. The work contributes a new system and technique to measure situation awareness that works for both manual and automated driving, and in laboratory and on-road testing. It is the only technique that leverages drivers' familiarity with current navigation tools in a naturalistic setting, thus offering high ecological validity, and addressing researchers' need for assessment across contexts.

This paper recounts our design process, including the Daze technique's user interaction model and implementation. It therefore emphasizes the contexts we are designing for, the design requirements we identified, the decisions we made, the validation performed, and our subsequent design reflections. In the following sections, we (1) detail the background and motivation for the work, (2) provide an overview of the design of Daze and describe how it works, (3) present an example of data collection and analysis using Daze, and (4) discuss the system in greater depth, to articulate different ways it can be adapted for related automotive studies.

SITUATION AWARENESS

Situation awareness is a critical factor in a driver's ability to make decisions to avoid hazards, plan routes and maintain safe travel. With the advent of automation, many of the current techniques used to assess driver performance [26], workload [13, 37] or behavior [27, 36] become useful only after a transition from autonomous to manual driving has occurred [14]. However, situation awareness is also important to assess prior to transition, given concerns over driver performance [11]. It can also be an indicator of a person's trust in automation [24], and hence will be an increasingly important measure in future automotive studies.

Awareness and Automation

In partially automated vehicles, drivers shift to a more supervisory role, however, they must remain situation aware as they may be called upon to make decisions or retake control of the vehicle if the automated driving system is unable to successfully address a situation [30]. To this point, the National Highway Traffic Safety Administration's (NHTSA) taxonomy of automated vehicles, levels 2 and 3 specify that the driver must be situation aware at all times (level 2), or can become situation aware and able to control the vehicle after some (brief) period (level 3) [21].

Maintaining sufficient alertness may be a challenge in automated vehicles, because with low attentional demand, cognitive resources may throttle back to conserve energy, making sudden rises in cognitive demand exceedingly challenging and potentially dangerous [36]. To quote Hancock's Automation Paradox, *"if you build systems where people are rarely required to respond, they will rarely respond when required"* [12]. Driving research has shown that rapid transitions from a low alertness state to active vehicle control can be quite dangerous [5, 19], as a lack of alertness and a lack of situation awareness can, when combined, lead to poor performance in a safety critical situation.

One of the most widely used models of situation awareness is Endsley's construct, which is comprised of three levels: (1) perception of elements in the environment, (2) comprehension of the situation, and (3) projection of the future state of the ambient environment [9, 10]. These three components are followed by decision and action components. Performance measures based on that action, such as controlling a vehicle, encapsulate the entire situation awareness process.

Techniques for Measuring Awareness

Current techniques for measuring situation awareness in driving settings were originally adapted from aviation testing. These methods include eye-tracking, the simulation freeze technique [9], subjective ratings, question probes (during and post activity), and task performance measures. Of these, measures that can assess awareness independent of driving performance are particularly valuable because they remove the role that driver experience, competence and familiarity with the vehicle play.

Eye-tracking

Eye-tracking can be an effective measure of level 1 situation awareness, providing an objective measure of what drivers see on the road, and thus, what they should perceive [32]. However, laying eyes on an object in the environment does not guarantee perception, as evidenced by the work of Chabris and Simons [6], and Drew et al. [7]. Eye-tracking research is relatively difficult, time consuming and expensive. It requires specialized equipment, and is highly sensitive to ambient conditions. In indoor simulators, eye-tracking systems are challenged to resolve targets located in multiple planes, and in outdoor environments, lighting conditions can interfere with tracking ability.

Physiological Measures

Using attentional resources for situation awareness increases cognitive workload, and the markers of neural activity can be detected using psychophysiological measures [5]. The deeper situation awareness levels of perception and projection are harder to measure: increased mental workload due to attentional processes can be detected by changes in blood flow, as indicated by changes in heart rate and heart rate variability [5]. Direct measures of brain activity can detect which areas of the brain are active, providing finer grained measures of what processes are at work: measuring neuroelectrical activity with an electroencephalogram (EEG) [4] and measurement of blood flow using functional near infrared spectroscopy (fNIRS) [29] have been used for driving research. These measures can be difficult for researchers to implement and interpret [33], especially in ecological settings such as in on-road studies, as they are highly susceptible to noise from muscle activity and from interference from automotive systems.

SAGAT

The freeze frame technique (Situation Awareness Global Assessment Technique, or SAGAT) involves halting a simulation in progress, and querying a person about activity in the environment, such as the position, type and future status of elements within the scene [9]. SAGAT is one of the most well-tested situation awareness measurement techniques for use in simulation environments. It was originally developed for flight interfaces, but has since been applied to driving tasks [16]. However, the freeze frame technique is limited to use within a simulator, and thus is not applicable for on-road driving settings. In addition, its intermittent halts are likely to decrease participants' sense of presence [15] of the simulation, and thus immersion in the situation, threatening a study's ecological validity.

SART
Subjective ranking techniques include ranking by either drivers themselves or by observers. In the area of military aviation, the SART technique involves pilots ranking themselves across 10 dimensions related to situation awareness post flight [35]. Although a well-tested technique, it provides only a subjective measure of driver situation awareness, and must be modified to be sufficiently applicable to non-experts in a driving context. Subjective rankings by observers can provide unobtrusive measures of situation awareness, and have the advantage of being used during live action evaluation or for post-hoc video coding. However, this approach requires many subject matter experts to review participants' behaviors, and results can still be of uncertain accuracy and reliability [28].

Question probes
Question probes can provide objective measures of elements perceived in the environment, and can also address level 2 and level 3 situation awareness. Current best practices for real-time probes are seen in the Situation Present Awareness Method (SPAM) [8]. This method utilizes response accuracy and time measures of situation awareness, although it has yet to be adapted for driving tasks.

Real-time probes are less intrusive than the freeze frame technique, and can also be used in real-world environments. However, they must be carefully designed so as not to draw the driver's attention to elements of the environment, nor should they be too cumbersome to address, as they can distract from drivers' abilities to perform their primary task. Post-driving probes do not interrupt the driving activity, but do rely on drivers' recall abilities for reliable measurement [19].

With automated driving, question probes become a much more useful technique, as the driver can respond to queries while the system controls the vehicle. This ability to respond in-the-moment improves ecological validity for studies in a simulator, as well as safety in an on-road vehicle, such as [2].

THE DAZE PLATFORM
Our intent with the Daze platform is to adapt current best practices in situation awareness assessment, so that they do not require physiological measures or retrospective assessments.

Design Requirements
Through our experience running participants through studies with both simulated and on-road scenarios, we have found that participant behavior often differs between the two. During simulations, drivers are willing to focus more of their attention on interfaces that we are developing than on the virtual environment; alternatively, during on-road studies, drivers focus their attention primarily on the road context, frequently looking away from the interface, due either to the vehicle's motion or concern over approaching danger.

Because simulation is usually used as a proxy for experiments that can be difficult, unsafe or expensive to run on the road, it is important to understand differences in behavior between simulator experiments and on-road experiments. To this end, a measurement technique for situation awareness that is usable in both environments is critical to making translational comparisons. We therefore developed a set of design requirements, where the new system should:

- Function in both simulation and on-road driving scenarios
- Not require halting the simulation or on-road driving
- Capture timing and response data based on driver responses
- Allow the driver to answer queries with a simple and unobtrusive graphical interaction

Regarding driver interaction, measurement techniques such as live question probes are typically verbal exchanges, where representing spatial data can be a challenge. The system should therefore leverage familiar and intuitive pointing gestures and graphical (map) views.

To be consistent with drivers' experiences, and accessible to the research community, the system should also:

- Reflect current (and thus, already familiar) navigation application interaction models
- Require minimal modification to a simulator's operation or a road vehicle's physical interface

With this perspective in mind, we have been developing and testing a driving specific situation awareness assessment technique to complement current evaluation techniques for automotive interfaces and systems.

Design Model
The platform is modeled after the Waze navigation application for mobile devices. A Waze user typically views a map that includes nearby users' locations, and can view, confirm, or post new alerts for the community about ongoing or transient driving conditions or hazards [31]. For example, if there is an accident in a highway exit lane, the application will warn approaching drivers that there is trouble ahead. It will also raise an alert requesting that the user confirm the ongoing existence of the disruption or any traffic that it has caused. If a driver notices that the roadway has cleared, then he or she can update the current state of the environment for the benefit of fellow drivers.

We observed that when drivers create and respond to Waze alerts, they are demonstrating levels 1 and 2 of Endsley's model—perception and comprehension of recent events, and possibly level 3—projection of the environment ahead. By modeling itself after Waze, Daze has a sensible, ecologically valid cover story for asking drivers' inputs about on-road conditions, which can then be used as measures of situation awareness. For example, Daze permits comprehension or projection questions such as *"is the traffic improving or worsening,"* or *"will a police car reach the accident within a minute?"*

Implementation
Since the application runs on a tablet or smartphone, it can be attached to a simulator, fixtured within a moving vehicle or held in someone's hand. During ordinary driving, it displays an overhead map of the environment centered on the car's current location, similar to other navigation tools. The map

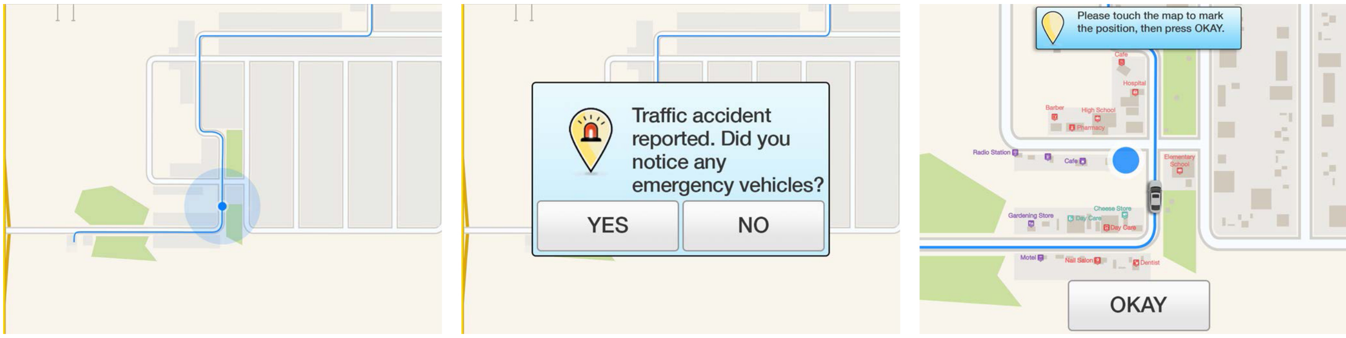

Figure 2. Daze ordinarily shows an overhead map centered on the car's location (left). After passing an event, it raises an alert (center). If the user confirms noticing the event, the application shows a closeup map view, asking the user to tap the screen to locate the event (shown with a blue dot), after which the user confirms the location (right).

tracks the car's movement so that the icon for the driver's vehicle is always at its center. Similar to typical navigation maps, it can show route overlays and annotations, and switch between alternate map views (road, satellite).

Maps

From the driver's perspective, the system appears similar in the simulator and on the road, but internally, there are several differences. In the simulator version of Daze, the map(s) shown were drawn using illustration software to represent the virtual environment, and are visually stylized to mimic a blend of Apple Maps[1] and Waze.[2] The simulator servers send current position and alert event messages, which the application reflects as updates to the live map. In the on-road version of Daze, the application builds upon Apple's MapKit framework to update map images as the vehicle moves and to trigger alert messages.

Views

Regarding the type of view to present, according to Taylor et al. [34], an egocentric view (driver's perspective) supports driving through local guidance, which is better for turn-by-turn directions, while an allocentric (overhead) view favors navigation requiring global awareness, including route learning. We chose the latter because Daze's questions ask participants to confirm and locate prior or anticipated events, which requires awareness of the evolving ambient environment.

Alerts

Shortly following events relevant to assessing situation awareness, Daze presents visible and (if needed) audible *alert* messages asking if the driver noticed the event. Alerts are styled after events typically called out in Waze, and that can easily be found on-road and replicated in a simulator, such as traffic, accidents or roadside construction. The alert screen includes *"yes"* and *"no"* buttons. If the driver selects no, the alert disappears; if he or she replies yes, a closeup view of the map, or *callout*, appears and asks the driver to tap the screen to identify the event's location (Fig. 2). These responses are all recorded and time-stamped.

[1] Apple Maps. 2016. http://www.apple.com/ios/maps/
[2] Waze. 2016. http://www.waze.com/

Alerts typically follow their corresponding events by a set time or distance, depending on the study's design, such that drivers cannot view events in side or rearview mirrors to confirm their existence or location. This prevents alerts from making drivers aware of events, and thus contaminating the awareness the technique is designed to assess.

We can use both the driver's error rate in detecting actual or false events, and also the location of the events as measures of situation awareness. Response times to the event queries are another measure of situation awareness, with lower times correlating with higher awareness [3]. Response time could also be a measure of distraction [18], or cognitive load, depending on the study. Additionally, the pop- up nature of alerts is similar to secondary task cognitive load testing techniques [23], and could be employed accordingly.

Both alert and callout screens dismiss themselves after 15 seconds, to avoid one alert cycle running into the next. We assume that a driver's lack of response to an alert or callout screen indicates that the driver has (likely) not noticed the alert.

In the simulator, Daze receives messages about real or false events via UDP messages sent over WiFi from a server. On the road, alerts are triggered by either (a) geofencing, which occurs when the device enters or exits a radius around a preset GPS coordinate, such as a known construction site or school, or (b) receiving UDP messages following the same messaging protocol, sent in-the-moment by an experimenter. Experimenter triggering of alerts while on the road can help researchers assess drivers' awareness of ephemeral events, such as traffic accidents. It can also be used during Wizard of Oz-style experimentation in the simulator or on the road [2, 20].

False alerts represent events which may never have existed, or which may have occurred earlier, but cleared over time. This capability is important both to better represent the way that Daze addresses situation awareness—probing drivers about events which may no longer exist—and to help break the pattern of *event occurs, alert appears,* that drivers might otherwise perceive.

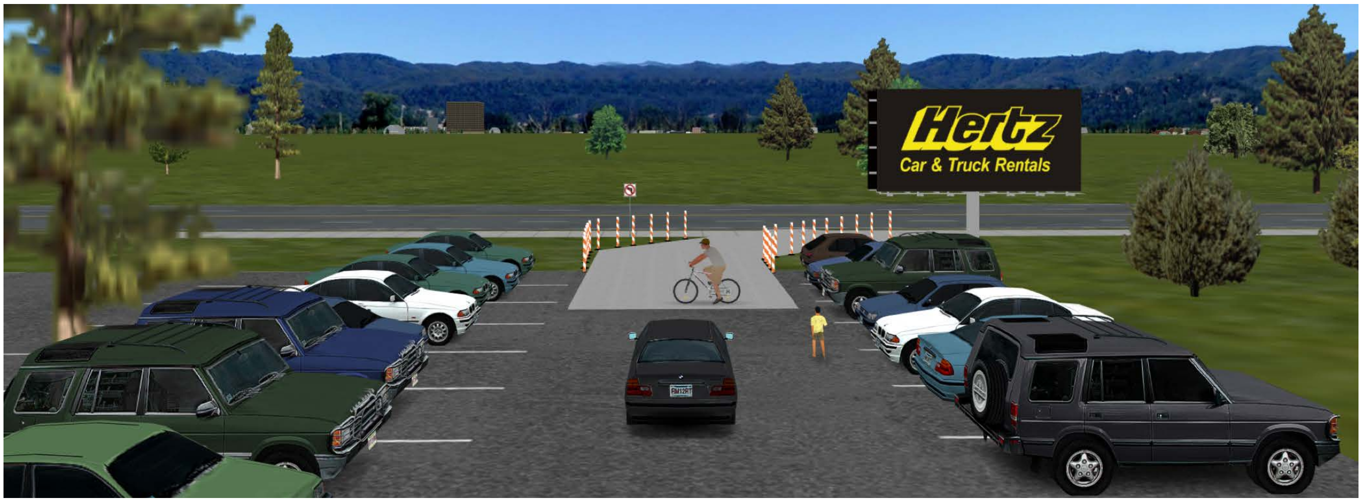

Figure 3. The parking lot where participants first view the test deployment's virtual environment. The child walks forward a few moments after the test begins. Just after, the bicyclist rides from right to left, accompanied by a ringing bell, which pans across the scene.

TEST DEPLOYMENTS

We conducted tests to check the Daze technique's ability to assess situation awareness. The primary goals are to (a) inform the design of the system's interaction model, and (b) provide data to develop an initial set of measures. Their context focuses on drivers' abilities to notice and localize events exterior to the car, and on how visible versus audible cues influence their awareness. These test deployments are the initial stage of a larger project concerning awareness of auditory events in autonomous driving. We present them here to illustrate how Daze can be used, and how the data that it provides can be analyzed, rather than as studies to be assessed on their own. Along the way, we discuss the technique's usability, and include preliminary findings about participants' awareness of audible versus visible events.

Participants

We recruited 12 participants, whose ages ranged from 17 to 58 years (M=28.2, SD=11.5), with 7 men and 5 women. Regarding driving habits, their reported years of driving experience ranged from 1 to 42 years (M=9.9, SD=11.2), with 83.3% having driven within the last two days, and 75.0% typically driving at least once per week. We offered a $25 gift card for participation.

Simulator

We conducted the test deployments in a high presence, whole-chassis, fixed base driving simulator. Participants sat in a late-model Toyota Avalon, and experienced the simulated driving course from the driver's seat. The simulator's visuals feature a 270°wraparound projection screen, rear projection screen (for the rear-view mirror), and two LCD side-view mirrors. The audio system includes two 6-channel speaker rings: one at ground level and another about four meters above the car, and a five-channel in-car audio system, used to enhance exterior sounds. A sub-woofer, attached to the car's body, simulates vibrations from an engine and from road roughness.

Events

A central concern in automation is *out-of-the-loop* experiences, where drivers may be focused on tasks other than driving, yet may need to intervene or resume control if the autonomous system fails to detect or handle potential hazards. To effectively assess situation awareness within this context, it is important that events reflect these potential hazards. We therefore adapt and include events from Bolstad et al.'s *goal-directed task analysis*, specifically, from the goal for avoiding road hazards outlined in [3].

The information requirements for the Bolstad's goal to *avoid road hazards* includes the presence and knowledge of pedestrians, school and construction zones, and road conditions. These potential hazards, in turn, include specific events, such as crosswalks, signs, lane changes, road equipment, speed limit changes, and the behavior of other drivers. These goals, decisions and information requirements were elicited from Bolstad's process, as described in [3].

Test Setting

Within the simulator, participants follow a course through a suburban-urban environment, which includes local, city and highway roads, during a fictional 20-minute autonomous morning drive to work. During a pre-test briefing, we ask participants to *"observe the vehicle's driving, and the environment around them."* These instructions are based on guidance on the driver's expected role, as outlined in the National Highway Safety Administration's recommendations for automation levels 3-4 [3].

The car navigates its way through the course under fully autonomous control, including negotiating its way out of, and into, parking spots at the test's start and completion (Fig. 3). The experience therefore requires no participant interaction with the car's pedals, shift lever or steering wheel. The steering wheel does turn, however, under its own power, in keeping with the car's maneuvers.

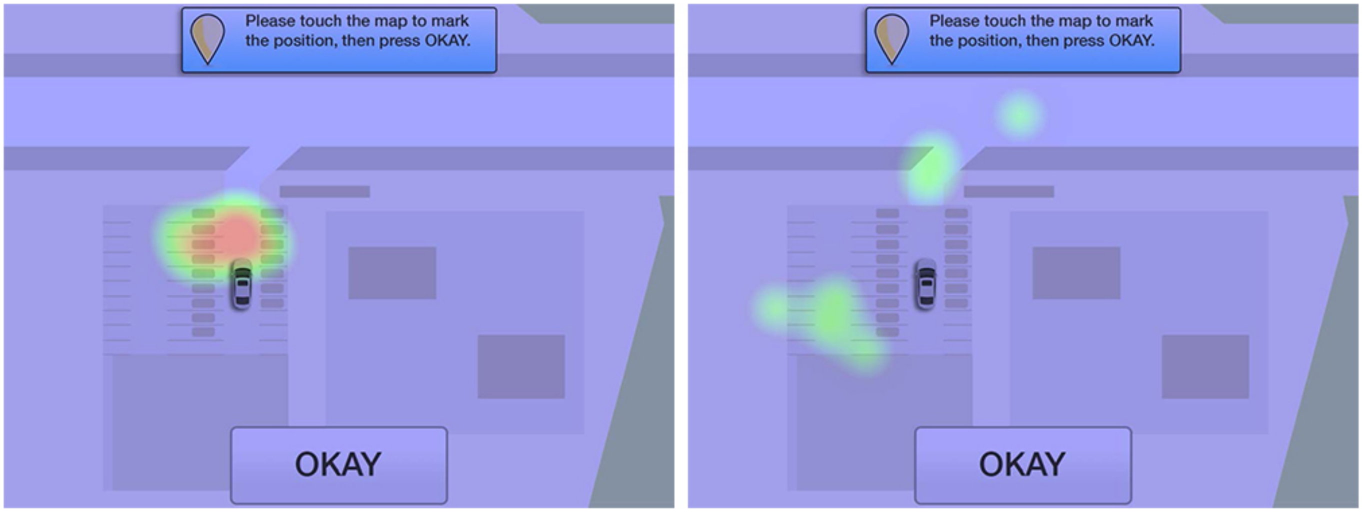

Figure 4. Heat maps showing participant responses to the location of the audible and visible bicyclist (left), and audible but not visible garbage truck (right), which both occur in the parking lot, moments apart, at the start of the drive.

Along the course, 11 events are presented, which are visible, audible, or both visible and audible. Events are based on the goal to avoid road hazards, and include: a child walking nearby, a bicyclist whizzing in front of the car while ringing a bell, a dog barking, an unseen truck unloading, a pedestrian crossing the street with an audible crosswalk signal, a pedestrian crossing another street with no audible signal, an emergency vehicle siren, a helicopter flying overhead, a police car partially concealed behind a highway billboard, a passing train, and a construction site with jack hammers and loaders. (The helicopter is not a road hazard event: we have included it to assess drivers' abilities to project a situation into the future—specifically, a nearby hospital as its destination.)

We attached a tablet computer running Daze to the car's center stack (Fig. 1). The program is set to raise alerts anywhere from 3 to 35 seconds after each event. Alerts are timed so that the event is no longer visible directly or in the mirror. Each Daze alert is accompanied by an audible tone (one of the tablet's built-in chimes). In addition to the alerts that match events in the environment, the tablet also presents three false alerts. These false events ask whether the driver notices a car honking its horn, the presence of a roadside hitchhiker, and a construction site, none of which exist in the simulation.

There are two alternative ways to present alerts that maintain consistency: appearing either a fixed time, or a fixed distance, after passing events. The fixed time approach lowers the interval before questions of recall, but due to variations in vehicle speed and visual obstructions, may allow participants to observe prior events in the surroundings or rearview mirror. We chose the fixed distance approach, as we are concerned with assessing awareness of events that may or may not have occurred, rather than the visibility of those events, which itself would influence situation awareness. For studies that focus on measuring near term comprehension or projection, a brief interval may be preferable.

The Daze application can be set to demonstrate—as the experimenter swipes the tablet's screen—what alerts and callouts look (and sound) like, and how participants may respond. This introductory demonstration is included during the initial pre-test briefing, and only occurs once.

Results

Participants correctly identified events that *did* occur 83.3% of the time, and rejected events that *did not* occur 91.7% of the time. Given our test environment, which includes events that are partially obscured visually, as might be the case on real-world roads, participants were most correct in identifying events that were both visible and audible, less so for events that were audible only, and notably less still for events that were visible only (Table 1).

Table 1. Participants' percentage correctness in identifying actual (visible, audible or both) and false events.

Visible Only	*Audible Only*	*Visible + Audible*	*Neither (False)*
58.3%	83.3%	95.8%	91.7%

Participants also identified the *locations* of events with greater precision and accuracy when those events were visible and audible, compared to only visible or only audible. Fig. 4 shows location responses for the bicyclist, who is seen riding across the screen and heard ringing a bell in front of the car, versus the garbage truck, which is heard nearby only moments later, but is not visible at all.

Regarding the time taken between the appearance of an alert and the driver's response, and the appearance of the following location callout and the driver's response, we found that participants were notably quicker for events that were both visible and audible (Table 2).

Table 2. Number of seconds for participants to respond to alerts and callouts for events that are visible, audible or both.

	Visible Only	*Audible Only*	*Visible + Audible*
Alert	4.5	4.2	2.8
Callout	7.9	8.1	5.0

Regarding participants' abilities to project a situation into the future (Endsley's level 3), 73% correctly indicated the (overhead and not visible) helicopter's direction, while 36% correctly identified the nearby hospital as its destination.

Data Analysis

In addition to directly reporting correctly identified events, locations, and response timing, we can derive measures that further inform us about drivers' awareness of events.

Locating Events

We averaged the Euclidean distance of responses for each event e across event types t (visible, audible, both) as a measure of dispersion:

$$d(t) = avg_{e \in t}\sqrt{(x_e - \bar{x}_e)^2 + (y_e - \bar{y}_e)^2} \quad (1)$$

For our test deployments, we recorded dispersion in map units of 1024×768. Knowing the map's scale then allows for translation to simulator or on-road distances. We found that participants were most precise in locating events that were both visible and audible (indicated by lower dispersion), and least precise for events that were audible only (Table 3).

Table 3. Dispersion *d(t)*, and quadrant-level accuracy, of participants' location responses for events that are visible, audible or both.

	Visible Only	*Audible Only*	*Visible + Audible*
Dispersion	53.3	116.3	48.16
Accuracy	85.0%	44.2%	87.8%

Determining response accuracy is challenged by events that are in motion. One approach is to subdivide the map around the car and note the percentage of responses that fall within the regions that contain the event path. The size of these regions then represents the measure's tolerance. For our test deployments, we divided callout views into quadrants centered on the vehicle, and found that accuracy displayed the same pattern as dispersion (Table 3).

Another approach to determine accuracy would apply a similar equation to dispersion, but with distances measured from events' positions or paths, rather than from the mean of participant responses.

Response Timing

We expected that longer response times, as well as larger variation in location estimates, would represent greater uncertainty over event locations, and this is supported by a Pearson correlation coefficient of r=0.82 between the mean of response times across participants for each event, and the dispersion for that event.

Drivers encounter a repeated event in different feature settings through two pedestrian crosswalks, 2.5 minutes apart. The first includes an audible crossing signal and the second does not. All participants correctly identified both events, with slightly faster response times for the second, inaudible event (M=3.0, SD=1.7 vs. M=2.3, SD=1.2). This counters the overall trend of longer response times for inaudible events, and may be due to the proximity (and thus visibility) of the event, combined with learning from repetition of the alert.

Discussion

Our primary objective for these test deployments is to demonstrate, through example, how to analyze the data that Daze collects to draw inferences about situation awareness. The data for these tests suggest that participants are better at *detecting* events around them that are audible, and better at *localizing* events that are visible. This agrees with intuition, and, we expect that audible events prompt drivers to perform a visual search, which, if successful, provides strong support for an event having occurred.

We did not expect strictly audible events to be more correctly identified and located than strictly visible events, although there are several possible explanations. First, audio cues are noticeable even if drivers' visual attention is focused elsewhere in the scene, or car, whereas visual cues can only be identified by direct or peripheral line of sight. Second, most of our events are in motion, relative both to the environment and to the moving car, so correctly identified locations could have been anywhere along those motion paths. This would have affected both visual and audible events, and either one could have been more strongly influenced. A likely outcome is that participants focused greater attention on confirming whether events had occurred at all, rather than locating them precisely.

Participant responses to the tablet, its alerts and location callouts were cautious, yet positive. When asked whether they agreed that the interaction with the application made them more aware of what was happening around them, 58.3% agreed. When asked whether the application interaction broke their sense of immersion in the environment, 66.7% disagreed.

During a semi-structured interview following each simulated drive, participants commented on circumstances where Daze would be most helpful to them: *"If someone's crossing in front of you, like a bicycle or something, I guess I would like it to notify me of events if it's going to affect my experience in the car. Like if the car's going to slam on the brakes."* And regarding awareness while driving: *"I'm not super observant to my environment, so I think that having some type of a cue to alert me to events is useful."* Responses reflect participants' perceptions of Daze as an awareness support technique—likely due to its similarity to familiar navigation applications—rather than an assessment technique, which contributes to ecological validity.

Participants' perceptions of their own abilities to localize audio cues is somewhat lower than the observational data suggests. For example: *"Oftentimes I wasn't exactly sure of what locations things that I was hearing were in. I heard them, and I heard them moving, but when I was asked to identify it, I wasn't about to pick it out."* And similarly: *"I heard more* *than I saw. There were instances I had to guess where the sound was coming from. I missed a few things too."*

ON-ROAD TESTING

We also tested the on-road version of Daze on public roads. Our goals were to (1) test the reliability of triggering events, (2) evaluate usability while on the road, and (3) gain experience designing a test course.

The first trial explored an automated driving scenario. The test vehicle was based on an on-road simulated autonomous driving platform (in the manner of [2]), and was implemented in a right-hand drive SUV. A passenger in the left front seat (the driver's position in the U.S.) held a tablet computer running the Daze application, while a Driving Wizard operated the vehicle. An Interaction Wizard sat in the rear seat and triggered alerts from a laptop that sent UDP messages over an ad-hoc wireless network to the tablet. The Interaction Wizard had a menu of pre-programmed alerts, including the presence of a road hazard, construction zone, accident, police car and nearby traffic. The Wizard observed the environment around the vehicle and raised event alerts after passing incidents that matched any of those on the menu.

The second trial explored a manual driving scenario. The test vehicle was a standard sedan, with a tablet computer running the Daze application mounted on the center stack (Fig. 1). In this case, the driver operated the vehicle, so there was no Driving Wizard. The Interaction Wizard triggered alerts from a remote location over the cellular data network, similar to the procedure above, only observing the environment through live video from a forward-facing camera on the car's dashboard. We limited use during manual driving due to concern over distraction, but our testing found it acceptable to issue alerts when the vehicle was stopped—as in traffic, at a sign or light, with 30 seconds available for alert and response.

The passenger in the first trial, and driver in the second trial, were both able to view and respond to alerts while the vehicles were en route, and the application collected the same type of response data as in the simulator tests. Since the Interaction Wizard triggered events opportunistically, and distances to events varied, the data were not directly comparable to that collected in the simulator.

We initially raised alerts without an audible cue, and found that the passenger missed several alerts, which timed out without a response. For studies where the response to the alert event is most important, we therefore recommend playing an accompanying audible cue. For studies where driver workload or distraction are the focus, the lack of an audible cue makes noticing the visible-only alert a measure in itself.

Events were not always perceived as we expected. For example, we passed a police station with a number of police vehicles parked outside, but the passenger did not acknowledge them in response to an alert. When asked about this after the drive, the passenger replied that she was expecting queries about police cars to be about those directly on the road, or stopped on the shoulder, not those in a parking lot somewhere to the side of the road, where it would be less relevant to someone driving.

With respect to technical implementation, the wireless connection and messaging worked as expected, with little lag and few dropped messages.

Some of Daze's current features come from discoveries made during this early on-road testing. For instance, we realized that certain landmark events that are always present—such as hospitals or school crossings—could be triggered by geofencing. This reduces the Interaction Wizard's tasks, allowing him or her to better attend to triggering more ephemeral events, such as the presence of pedestrians or bicyclists, emergency vehicles or traffic. We also found that a construction site, which was present during course planning, had cleared before the test drive. We were still able to trigger a construction alert at a subsequent site, highlighting both the inherent variability involved in running tests on real roadways, and also the usefulness of being able to trigger events in an ad-hoc basis.

Overall, on-road testing verified that live event triggering is a feasible approach, and it unearthed several challenges facing on-road usage compared to a simulated environment.

IMPLEMENTATION NOTES

Test Deployments

In our test deployments, we set out to recreate conditions that experimenters might face in simulated and, especially, real-world settings, including circumstances that could pose a challenge to clean data collection, such as fully or partially obscured, or moving, objects of interest. Findings demonstrate that the Daze application, the data it records, and analysis of that data, are able to characterize drivers' awareness of actual versus false event occurrences and their locations. By examining the timing of responses, we can also infer, to some degree, drivers' certainty of their responses, without directly asking.

The test deployments serve primarily as an exploration of the types of data that can be collected, and analysis that can be performed, using the Daze technique to better understand situation awareness. Our findings are influenced by the settings that we created, and questions that we asked about them, and results should be taken as such. For example, halfway through the drive, a pedestrian walks through a crosswalk at the intersection just ahead of the car, and we later ask the participant to locate that pedestrian on a map. Since the pedestrian was seen traveling along a path, anywhere across that path could have been a reasonable response. We therefore focused our analysis on response dispersion, and how that relates to response timing, rather than the exact position or direction implied by particular responses. Another approach might be to clarify the questions posed in alert and callout messages to account for variability in event paths or timing—for example, eliciting more or less specificity in their responses.

On-Road Testing

Insights from the on-road testing include that the platform performed as we had planned, but that field conditions were less predictable than we had expected. As a takeaway, we expect that triggering alerts through geolocation alone may be insufficient, due to the transience of people and places in the physical world, and that the capability to support improvisational action by researchers can be most valuable.

One limitation to this method of on-road situation awareness assessment is that tested events are limited to those that the Interaction Wizard is aware of, or those that have been previously been programmed into a simulator course. This suits the controlled experiments that the technique is designed for, but does not capture the participants' general awareness of all interesting things that might happen in a real environment. It is possible to extend the questions that are asked by incorporating prior information about the course through maps or pre-test scouting drives, or by employing additional experimenters in the field or confederate vehicles that send messages to the Interaction Wizard. Other post-facto analysis techniques, such as interactive video review or post-drive questionnaires, could be used in conjunction with the Daze technique to address such events.

Questioning and Awareness

Daze shares limitations presented by all concurrent situation awareness assessment techniques, including SAGAT and SPAM. For example, in asking people about events, we may bias them towards noticing events that they might otherwise not have noticed. And yet, in their 2013 evaluation of validity and intrusiveness of real-time questioning, Bacon and Strybel *"failed to find evidence supporting the idea that online probing affects subsequent situation awareness"* [1].

We try to limit potential influence by being careful to incorporate *false positive* event questions about things that are not present. Also, we wait several seconds to ask about an event after it has occurred, to make sure that drivers cannot retroactively check behind them for the event in question. In addition, we follow procedures suggested by Endsley [10], such as careful experimenter instruction and randomizing query topic and timing to minimize influence. Endsley reports that *"no intrusion on performance has been repeated in numerous other studies"* using these procedures.

Other limitations need to be addressed through study design. For example, since it is the difference in measurement across conditions rather than the absolute measurement that matters, some bias in events (such as their size, location, apparent motion, or sound profile) can be alright so long as they apply evenly across all conditions. It is also a good idea to include events that induce several senses, including front-rear or side-side movement, high or low frequency vibration, or even scent, so as to assess awareness of conditions and factors that are not exclusively visible or audible.

Questioning and Arousal

Regarding concern over drivers maintaining situation awareness in conditions of low arousal, Pierce et al. find no difference in self perception of workload between real-time and post-study questioning techniques, although they do find that real-time questioning is comparable to other secondary tasks [25]. This suggests that exploring lower levels of situation awareness using Daze may be possible, within the constraints of (any) real-time assessment noted above.

In early testing in the simulator and on-road, we found that participants focus their visual attention on the surroundings, or alternatively, close their eyes to rest, and therefore miss many alerts when they appear. Because we focused on assessing awareness of visible and audible events, and identifying actual versus false events, we set the application to issue a tone with each alert. For research agendas with a stronger emphasis on driver distraction, focus of attention, or awareness of specific events or incoming alerts, the accompanying tone can be disabled through a setting.

Maps and Orientation

The callout screen, which asks participants to locate events on a close-up map, initially included only a map image. We received feedback, during initial testing, that drivers found it challenging to identify their own location, let alone that of another element in the scene, relative to themselves. Thus, we added an overhead image of the car, at a slightly exaggerated scale (similar to Waze), in the center of each map, which improved drivers' responsiveness.

We positioned the image of the car at the location that the car was in the past, when the event occurred; however, it is reasonable to position the car at its location in the present, at the moment when the alert is raised, or to have it reflect the car's continued motion. Indicating which is the case to participants prior to the drive will clarify the procedure, and likely improve responses. We intend to explore these differences during upcoming studies.

REFLECTIONS AND FUTURE WORK

We have shown that Daze collects data useful for the measurement of situation awareness, demonstrated an example analysis using that data, and described many of its design considerations and nuances.

Our test deployments evaluated the platform primarily taking a reflective frame: that is, investigating situation awareness through recall of people, places and events previously observed in the environment. While it is important to understand driver awareness in an automated driving context, the technique can also investigate cuing and information delivery, or provide information about future situations. The test environment included one event that reflects the ability to project future state: the audible, but not visible, helicopter flying toward the hospital, which about one-third of participants inferred as its destination. Daze can thus be used as a means to test the anticipatory abilities of a driver acting in a supervisory role, in addition to other methods, such as driving performance measurement.

We are currently running simulator and on-road studies that employ the technique to explore how alternative interface designs and interaction patterns affect situation awareness. We envision alternative versions, which focus on navigation, text messaging, and even gaming.

Separating the measurement of situation awareness from driving performance, as is necessary in autonomous driving situations, makes the Daze technique useful in other awareness-dependent contexts. We imagine applying a similar approach to personal navigation, augmented reality or wearable computing, to understand how applications or equipment affect users' awareness of the world around them.

The Daze source code and executable are available at `https://github.com/CDR-IxD/Daze`.

ACKNOWLEDGMENTS

The authors thank Sile O'Modhrain for contributing to test deployment procedure and audio and design and Dylan Moore for assistance with data and statistical analyses.

This research was supported by Stanford Center for Design Research's Affiliate Program, and conducted under Stanford University IRB protocols 30016 and 32402.

REFERENCES

1. L. Paige Bacon and Thomas Strybel. 2013. Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers. *Safety Science* 56 (2013), 89–95. http://dx.doi.org/10.1016/j.ssci.2012.06.019
2. Sonia Baltodano, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju. 2015. The RRADS platform: A real road autonomous driving simulation. In *Proc. of the ACM/IEEE Intl. Conf. on Automotive User Interfaces and Interactive Vehicular Interactions*. ACM, 281–288. http://dl.acm.org/citation.cfm?id=2799250.2799288
3. Cheryl A. Bolstad, Haydee M. Cuevas, Jingjing Wang-Costello, Mica R. Endsley, and Linda S. Angell. 2009. Measurement of situation awareness for automobile technologies of the future. In *Performance Metrics for Assessing Driver Distraction*. SAE, 195–213.
4. G. Borghini, G. Vecchiato, J. Toppi, L. Astolfi, A. Maglione, R. Isabella, C. Caltagirone, W. Kong, D. Wei, Z. Zhou, and L. Polidori. 2012. Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual Intl. Conf. of the IEEE*. IEEE, 6442–6445. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6347469
5. Karel A. Brookhuis, Dick De Waard, and Wiel H. Janssen. 2001. Behavioural impacts of advanced driver assistance systems—an overview. *European Journal of Transport and Infrastructure Research* 1, 3 (2001), 245–253. http://www.ejtir.tbm.tudelft.nl/issues/2001_03/pdf/2001_03_02.pdf
6. Christopher Chabris and Daniel Simons. 2011. *The invisible gorilla: And other ways our intuitions deceive us*. Broadway Books. http://dx.doi.org/10.1080/17439760.2012.753466
7. Trafton Drew, Melissa L. H. Võ, and Jeremy M. Wolfe. 2013. The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science* 24, 9 (2013), 1848–1853. https://doi.org/10.1177/0956797613479386
8. Frank T. Durso, Andrew R. Dattel, S. Banbury, and S. Tremblay. 2004. SPAM: The real-time assessment of SA, S. Banbury and S. Tremblay (Eds.). *A Cognitive Approach to Situation Awareness: Theory and Application* (2004), 137–154.
9. Mica R. Endsley. 1995a. Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 65–84. http://hfs.sagepub.com/content/37/1/65.short
10. Mica R. Endsley. 1995b. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 32–64. http://hfs.sagepub.com/content/37/1/32.short
11. Mica R. Endsley and David B. Kaber. 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics* 42, 3 (1999), 462–492. http://dx.doi.org/10.1080/001401399185595
12. P. A. Hancock. 2014. Automation: How much is too much? *Ergonomics* 57, 3 (2014), 449–454. http://dx.doi.org/10.1080/00140139.2013.816375
13. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183. http://dx.doi.org/10.1016/S0166-4115(08)62386-9
14. Key Jung Lee, Yeon Kyoung Joo, and Clifford Nass. 2014. Partially intelligent automobiles and driving experience at the moment of system transition. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. ACM, 3631–3634. http://dl.acm.org/citation.cfm?id=2557370
15. Kwan Min Lee. 2004. Presence, explicated. *Communication Theory* 14, 1 (2004), 27–50. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2885.2004.tb00302.x/full
16. R. Ma and David B. Kaber. 2007. Situation awareness and driving performance in a simulated navigation task. *Ergonomics* 50, 8 (2007), 1351–1364. http://dx.doi.org/10.1080/00140130701318913
17. Nikolas Martelaro, David Sirkin, and Wendy Ju. 2015. Daze: A real-time situation awareness measurement tool for driving. In *Proc. of the ACM/IEEE Intl. Conf. on Automotive User Interfaces and Interactive Vehicular Interactions*. ACM, 158–163. http://dl.acm.org/citation.cfm?doid=2809730.2809753
18. M. H. Martens and W. Van Winsum. 2000. Measuring distraction: the peripheral detection task. *TNO Human Factors, Soesterberg, Netherlands* (2000). http://dx.doi.org/10.1016/j.trf.2005.04.009

19. Brian Ka-Jun Mok, Mishel Johns, Key Jung Lee, Hillary Page Ive, David Miller, and Wendy Ju. 2015a. Timing of unstructured transitions of control in automated driving. In *Intelligent Vehicles Symposium*. IEEE, 1167–1172. http://dx.doi.org/10.1109/IVS.2015.7225841

20. Brian Ka-Jun Mok, David Sirkin, Srinath Sibi, David Bryan Miller, and Wendy Ju. 2015b. Understanding driver-automated vehicle interactions through Wizard of Oz design improvisation. In *Proc. Intl. Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. 386–392.

21. National Highway Traffic Safety Administration and U.S. Department of Transportation. 2016. *Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety*. Washington, DC.

22. National Highway Traffic Safety Administration and others. 2013. *Preliminary Statement of Policy Concerning Automated Vehicles*. Washington, DC.

23. Fred Paas, Juhani E. Tuovinen, Huib Tabbers, and Pascal WM Van Gerven. 2003. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist* 38, 1 (2003), 63–71. http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3801_8

24. Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. 2008. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making* 2, 2 (2008), 140–160. http://edm.sagepub.com/content/2/2/140.short

25. Russell Pierce, Kim-Phuong L. Vu, Jimmy Nguyen, and Thomas Strybel. 2008. The relationship between SPAM, workload, and task performance on a simulated ATC task. In *Proc. of the Human Factors Ergonomics Society*. 34–38. https://doi.org/10.1177/154193120805200109

26. A. R. Quimby and G. R. Watts. 1981. *Human factors and driving performance*. Technical Report. http://trid.trb.org/view.aspx?id=174634

27. Thomas A. Ranney. 1994. Models of driving behavior: A review of their evolution. *Accident Analysis & Prevention* 26, 6 (1994), 733–750. http://www.sciencedirect.com/science/article/pii/0001457594900515

28. Paul Salmon, Neville Stanton, Guy Walker, and Damian Green. 2006. Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics* 37, 2 (2006), 225–238. http://www.sciencedirect.com/science/article/pii/S0003687005000554

29. Angelo Sassaroli, Feng Zheng, Leanne M. Hirshfield, Audrey Girouard, Erin Treacy Solovey, Robert JK Jacob, and Sergio Fantini. 2008. Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences* 1, 02 (2008), 227–237. http://www.worldscientific.com/doi/abs/10.1142/S1793545808000224

30. Thomas B. Sheridan. 2006. Supervisory control. *Handbook of Human Factors and Ergonomics* (2006), 1025–1052. http://onlinelibrary.wiley.com/doi/10.1002/0470048204.ch38/summary

31. Thiago H. Silva, Pedro O.S. Vaz de Melo, Aline Carneiro Viana, Jussara M. Almeida, Juliana Salles, and Antonio AF Loureiro. 2013. Traffic condition is more than colored lines on a map: Characterization of Waze alerts. In *Social Informatics*. Springer, 309–318. http://link.springer.com/chapter/10.1007/978-3-319-03260-3_27

32. M. W. Smolensky. 1993. Toward the physiological measurement of situation awareness: The case for eye movement measurements. In *Proc. of the Human Factors and Ergonomics Society*, Vol. 41.

33. Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M. Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert JK Jacob. 2009. Using fNIRS brain sensing in realistic HCI settings: Experiments and guidelines. In *Proc. ACM Symposium on User Interface Software and Technology*. ACM, 157–166. http://dl.acm.org/citation.cfm?id=1622207

34. Holly A. Taylor, Susan J. Naylor, and Nicholas A. Chechile. 1999. Goal-specific influences on the representation of spatial perspective. *Memory & Cognition* 27, 2 (1999), 309–319. http://link.springer.com/article/10.3758/BF03211414

35. R. M. Taylor. 1990. Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17* (1990).

36. Kristie L. Young, Michael A. Regan, and John D. Lee. 2009. *Measuring the effects of driver distraction: Direct driving performance methods and measures*. CRC Press, Boca Raton, FL. http://www.crcnetbase.com/doi/abs/10.1201/9781420007497.ch7

37. Mark S. Young and Neville A. Stanton. 2002. Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science* 3, 2 (2002), 178–194. http://www.tandfonline.com/doi/abs/10.1080/14639220210123789